

• Original Paper •

A Hybrid Statistical-Dynamical Downscaling of Air Temperature over Scandinavia Using the WRF Model

Jianfeng WANG^{*1}, Ricardo M. FONSECA², Kendall RUTLEDGE³,
Javier MARTÍN-TORRES^{2,4,5}, and Jun YU¹

¹*Department of Mathematics and Mathematical Statistics, Umeå University, SE 901 87, Umeå Sweden*

²*Group of Atmospheric Science, Division of Space Technology, Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology, SE 971 87 Luleå, Sweden*

³*Novia University of Applied Sciences, PO BOX 6, FI-65201 Vaasa, Finland*

⁴*Instituto Andaluz de Ciencias de la Tierra, 18100 Granada, Spain*

⁵*The Pheasant Memorial Laboratory for Geochemistry and Cosmochemistry, Institute for Planetary Materials, Okayama University at Misasa, Tottori 682-0193, Japan*

(Received 23 April 2019; revised 8 August 2019; accepted 26 August 2019)

ABSTRACT

An accurate simulation of air temperature at local scales is crucial for the vast majority of weather and climate applications. In this work, a hybrid statistical–dynamical downscaling method and a high-resolution dynamical-only downscaling method are applied to daily mean, minimum and maximum air temperatures to investigate the quality of local-scale estimates produced by downscaling. These two downscaling approaches are evaluated using station observation data obtained from the Finnish Meteorological Institute over a near-coastal region of western Finland. The dynamical downscaling is performed with the Weather Research and Forecasting (WRF) model, and the statistical downscaling method implemented is the Cumulative Distribution Function-transform (CDF-t). The CDF-t is trained using 20 years of WRF-downscaled Climate Forecast System Reanalysis data over the region at a 3-km spatial resolution for the central month of each season. The performance of the two methods is assessed qualitatively, by inspection of quantile-quantile plots, and quantitatively, through the Cramer-von Mises, mean absolute error, and root-mean-square error diagnostics. The hybrid approach is found to provide significantly more skillful forecasts of the observed daily mean and maximum air temperatures than those of the dynamical-only downscaling (for all seasons). The hybrid method proves to be less computationally expensive, and also to give more skillful temperature forecasts (at least for the Finnish near-coastal region).

Key words: WRF, air temperature, Cumulative Distribution Function-transform, hybrid statistical–dynamical downscaling, model evaluation, Scandinavian Peninsula

Citation: Wang, J., R. M. Fonseca, K. Rutledge, J. Martín-Torres, and J. Yu, 2020: A hybrid statistical–dynamical downscaling of air temperature over Scandinavia using the WRF model. *Adv. Atmos. Sci.*, **37**(1), 57–74, <https://doi.org/10.1007/s00376-019-9091-0>.

Article Highlights:

- Hybrid statistical–dynamical and dynamical-only downscaling techniques are assessed for air temperature forecasts over Scandinavia.
- WRF output at a 3-km spatial resolution combined with CDF-t is used as the hybrid technique.
- WRF output at a 1-km spatial resolution is used as the dynamical-only downscaling approach.
- The hybrid method is computationally cheaper and gives more skillful forecasts, particularly of daily mean and maximum air temperatures.

* Corresponding author: Jianfeng WANG
Email: jianfeng.wang@umu.se

1. Introduction

General circulation models (GCMs) often fail to simulate many of the processes that drive regional and local climate variability. This primarily results from their coarse spatial resolutions (typically 100–200 km), which do not allow for proper representations of the regional topography and subgrid-scale features. When evaluated against observational data, the raw GCM output data are often found to have systematic biases (Vrac et al., 2012; Vigaud et al., 2013), limiting their usefulness for direct downstream applications. The same is true for reanalysis datasets, obtained by combining GCM output with observational data (Dulière et al., 2011). There is thus a need to make more skillful local-scale predictions using the lower-resolution GCM or reanalysis data as the input, and this procedure is known as downscaling.

There are two main downscaling approaches: dynamical downscaling and statistical downscaling. In the former, a higher-resolution model, such as a regional climate model (RCM), is driven by a GCM or reanalysis data and run at spatial resolutions of up to a few meters (e.g., Aitken et al., 2014), at which complex topography and smaller-scale processes are better represented (Laprise, 2008). This approach can give a very good simulation of local atmospheric conditions (Pan et al., 2012; Soares et al., 2012; Warrach-Sagi et al., 2013; Aitken et al., 2014; Heikkilä et al., 2014). However, it has some drawbacks: e.g., it requires a significant amount of computational resource, particularly for very-high-resolution simulations, and the RCM performance is also strongly dependent on the GCM/reanalysis boundary forcing data (Fowler et al., 2007). Further, when forced with coarse-resolution data and run without any additional constraints, the interior fields of the RCM's domain can diverge substantially from the driving fields (e.g., Bowden et al., 2012, 2013). One way to prevent this is to relax or nudge the fields in the interior of the domain to those of the coarse-grid data used to force the model (e.g., Waldron et al., 1996; von Storch et al., 2000). However, doing so is not the meteorological community's consensus, as it removes freedom from the model's large scales (e.g., Miguez-Macho et al., 2004; Alexandru et al., 2009), even though it has proven to be effective (e.g., Otte et al., 2012). There are two main nudging techniques employed in RCMs: grid (or analysis) nudging and spectral nudging. In the former, the prognostic variables are relaxed towards a reference state at every grid point (e.g., Stauffer and Seaman, 1990), whereas in the latter only some zonal and meridional wavenumbers are retained, with all other waves filtered out (e.g., Miguez-Macho et al., 2004, 2005). Both techniques have been successfully applied in numerical simulations (e.g., Soares et al., 2012; Heikkilä et al., 2014; Steinhoff et al., 2014; Ma et al., 2016; Wootten et al., 2016).

A computationally less demanding and more flexible alternative to dynamical downscaling is statistical downscaling. Statistical methods have been used for some time to downscale and predict variables such as temperature and precipitation, and range from simple regression methods [includ-

ing multiple linear regression and partial least-squares regression (e.g., Ke et al., 2011; Fan et al., 2012)] to more complex techniques such as singular value decomposition (e.g., Landman and Tennant, 2000; Wei and Huang, 2010), artificial neural networks (e.g., Wilby and Wigley, 1997), and the method of analogs (e.g., Zorita and von Storch, 1999).

Statistical techniques are often used to correct GCM/RCM parameter field biases (Déqué, 2007; Pierce et al., 2015). They are typically applied in a two-step process: in step (1), known as the past or “training” period, the statistical relationships between local climate variables (e.g., from a weather station) and large-scale fields (e.g., from a numerical model) are developed; and in step (2), known as the future or “prediction” period, they are applied to some period of interest. For example, model data for the past climate can be used for “training” purposes with the statistical relationships then applied to the output of the same model for a future climate simulation, to infer the local climate change signal (Diaz-Nieto and Wilby, 2005). As discussed in Pierce et al. (2015), some of the most commonly employed statistical techniques make use of the cumulative distribution functions (CDFs) of modelled and observed data. A CDF, denoted as $F_X(x)$, is a distribution function of a random variable X evaluated at x that returns the probability of $X \leq x$. It varies between 0, at the minimum value of a CDF, and 1, at its maximum value. When two CDFs are plotted against each other, a quantile-quantile (Q-Q) plot is obtained (Field, 2013). A simple bias correction approach is quantile mapping. In this technique, the Q-Q plot for the “training” period, obtained using past model and observational data, is used to bias-correct the model forecasts for the “prediction” period. Another commonly used method is equidistant quantile matching. Here, and at a given quantile, the model-predicted change signal is added to the observed data for the past period. The advantage of this approach is that, if the model bias is invariant with time, it will be effectively removed when taking the difference of the future and past forecasts. The statistical method used in this work is the CDF-transform, or CDF-t (Michelangeli et al., 2009). In this technique, a transformation that projects the model's CDF to the observed CDF is computed over the past period and is subsequently applied to the model's future CDF. This statistical stationarity assumption (invariance of statistical moments in time) is commonly made in CDF-t studies, including this one, even though it has recently been challenged (Dixon et al., 2016).

The CDF-t method has been used in a wide range of studies. It has been employed to directly downscale reanalysis data (Kallache et al., 2011), and it has also been applied in conjunction with a dynamically downscaled (GCM/RCM) product [hybrid technique (e.g., Lavaysse et al., 2012; Vrac et al., 2012; Flaounas et al., 2013; Wu et al., 2018)]. The CDF-t approach has been found to be very effective at removing model biases (e.g., Famien et al., 2018). In fact, Colette et al. (2012) concluded that a hybrid downscaling technique, using the Weather Research and Forecasting (WRF) model (Skamarock et al., 2008) and the CDF-t approach, can give nearly unbiased, high-resolution, physically consistent,

three-dimensional fields that can be used for climate impact studies. Vigaud et al. (2013) applied the CDF-t to GCM surface temperature and precipitation predictions over southern India. The authors found a substantial improvement compared to the raw GCM data, regarding the distribution, seasonal cycle and monsoon daily means for different environments in the Indian subcontinent. Bechler et al. (2015) down-scaled WRF-predicted precipitation data over southern France using different hybrid downscaling approaches and found that CDF-t was one of the best performing statistical techniques. The successful implementation of the CDF-t for different meteorological studies, such as those mentioned above, justifies the choice of the technique for this work.

In this paper, the CDF-t technique is first developed using the 3-km output of the WRF model, WRF[3 km], and station data from the Finnish Meteorological Institute (FMI). For the central month of each season in a given year, the developed technique is used in combination with a 3-km WRF output, with the hybrid statistical–dynamical downscaling, WRF[3 km]+CDF-t, compared against a 1-km dynamically downscaled product, WRF[1 km], over Scandinavia. If WRF[3 km]+CDF-t is found to give more accurate predictions than WRF[1 km], it can be argued that there is no substantial added value in conducting numerical simulations over this region at kilometer and perhaps sub-kilometer resolutions. Should this be true, and given that a very-high-resolution dynamical downscaling is extremely computationally expensive, the findings of this work will provide guidance into the setup of future model runs over Scandinavia.

The focus of the work is on the air temperature as (i) it is a crucial field for the vast majority of weather and climate applications, including climate change studies (e.g., Overland et al., 2016), and (ii) this variable is projected to change significantly in the target region in a hypothetical warming climate (e.g., Jylha et al., 2004; Hanssen-Bauer et al., 2005). In addition, RCMs, including the WRF model, are found to have large temperature biases in the region (e.g., Kotlarski et al., 2014; Katragkou et al., 2015), mak-

ing it of interest to investigate how much of an added value a hybrid downscaling approach will give.

This paper is organized as follows. The models, datasets and diagnostics are introduced in section 2. The results are presented and discussed in section 3, while the main findings are outlined in section 4.

2. Model, datasets and diagnostics

2.1. WRF simulations

In this study, the WRF model is run over the Honkajoki wind farm in near-coastal western Finland with two different configurations, given in Fig. 1. For the “training” period, a 20-year run (April 1997 to March 2016) with 27-, 9- and 3-km grids is conducted, hereafter referred to as the WRF[3 km] simulation. For the “forecast” period, WRF is run for 4 months (April, July, October 2016 and January 2017) with an additional 1-km nest, WRF[1 km]. The computational cost of the WRF[1 km] simulations hinders us from taking full seasons as prediction periods, so we focus instead on the central month of each season. The rather short evaluation period is a limitation of this study, and a multi-year simulation would be needed given the significant interannual variability of the atmospheric conditions in northern Europe (e.g., Toniazzo and Scaife, 2006; Gray et al., 2016). The Honkajoki near-coastal region is selected given its strong seasonal contrast, typical of high-latitude climates (snow-covered terrain and coastal sea-ice in winter versus open water and snow-free land in summer). The presence of a wind park in this area provides a potential extension of the methodology developed here for wind-energy-related applications. Figure 1 shows the spatial extent of the model grids used in both simulations: 27-9-3 km grids in the former, and 27-9-3-1 km grids in the latter. The initial and boundary conditions for the runs are obtained from the National Centers for Environmental Prediction Climate Forecast System Reanalysis (CFSR; Saha et al., 2010) 6-hourly data (spatial resolu-

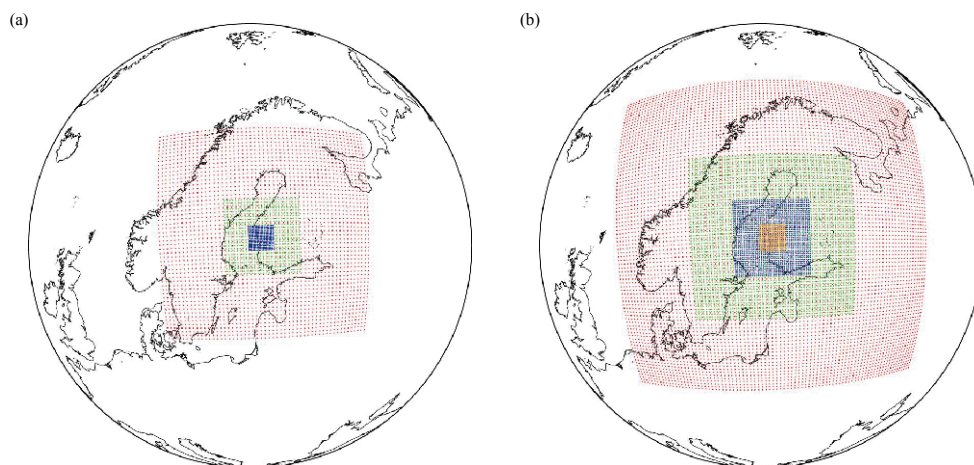


Fig. 1. Spatial extent of the model grids used in the (a) 20-year and (b) 1-year WRF simulations with the boundary regions excluded. The spatial resolutions of the grids are 27 km (red), 9 km (green), 3 km (blue) and 1 km (orange).

tion of $0.5^\circ \times 0.5^\circ$). CFSR has been shown to be one of the best performing reanalysis datasets in midlatitude environments (e.g., Ebisuzaki and Zhang, 2011; Free et al., 2016). It is important to note that the spatial resolution of the CFSR data is higher than that employed in the majority of GCMs used for climate simulations: e.g., the output of the Community Climate System Model, version 3 (Collins et al., 2006), frequently used to drive future regional climate change simulations, is available at a 1.4° spatial resolution (Bruyère et al., 2014). CFSR data are employed here to force WRF owing to their good performance over the region, and given that the focus of this work is on the assessment of the performance of the two downscaling methods. In order to minimize the accumulation of integration errors, and following Lo et al. (2008), the simulations are broken into several overlapping periods. For the 20-year WRF simulation, WRF[3 km], the model is run for 13 months at a time (March year Z to March year $Z+1$) with the first month regarded as model spin-up. The (re-)initialization takes place in spring so as to keep the summer (June to August) and winter (December to February) seasons within one continuous integration. The WRF[1 km] simulations are conducted separately for April 2016 (spring), July 2016 (summer), October 2016 (autumn) and January 2017 (winter), with a 1-day spin-up before each.

The setup of the WRF simulations presented here is the same as in Wang et al. (2019) and Duran et al. (2019), which is found to work well for this higher-latitude region. The following physics parameterizations are used: the Thompson double-moment 6-class cloud microphysics scheme (Thompson et al., 2008); the Rapid Radiative Transfer Model for Global models, for both shortwave and long-wave radiation (Iacono et al., 2008), with the Tegen et al. (1997) climatological aerosol distribution; Mellor–Yamada Nakanishi and Niino (MYNN) level 2.5 (Nakanishi and Niino, 2006) with Monin–Obukhov surface layer parameterization (Monin and Obukhov, 1954); the Fitch et al. (2012) wind farm parameterization scheme configured for the Honkajoki wind farm in western Finland; the four-layer Noah land surface model (Noah LSM; Chen and Dudhia, 2001); and the Betts–Miller–Janjić (BMJ) cumulus scheme (Janjić, 1994) coupled with the Koh and Fonseca (2016) precipitating convective cloud scheme. The BMJ scheme is switched off in the 3- and 1-km grids. The combination of the Thompson cloud microphysics scheme with the MYNN PBL scheme was found to give the best results for a dynamical downscaling over a wind farm in Sweden (Davis et al., 2014). An interactive prognostic scheme for the sea surface skin temperature based on Zeng and Beljaars (2005) is also employed, with the sea surface temperatures read in from CFSR every 6 h and linearly interpolated in time in order to have a continuous-varying forcing on the skin layer. Also read in from CFSR are the fractional sea-ice coverage and sea-ice depth, with the sea-ice albedo being a function of the air temperature, skin temperature and snow (Mills, 2011). Gravitational settling of cloud drops in the atmosphere is parameterized in the model as in Dwynerker

(1991) and Nakanishi (2000), while cloud water (fog) deposition onto the surface due to turbulent exchange and gravitational settling is treated using the fog deposition estimation scheme (Katata et al., 2011).

Grid (or analysis) nudging towards CFSR is employed in the outermost (27-km) domain with the water vapor mixing ratio nudged in the mid- and upper-troposphere and the horizontal wind components and potential temperature perturbation nudged in the upper-troposphere and stratosphere. The former is needed to correct the WRF humidity biases, as reported in Merino et al. (2015) for example, while the latter is justified by the reduced vertical resolution of WRF in the upper troposphere and stratosphere compared to the model used to generate the CFSR data (Saha et al., 2010), and the need to properly simulate the flow in the region given the strong stratosphere–troposphere coupling in the winter season (e.g., Kidston et al., 2015). The nudging time scale is set to 1 h for all fields. Employing analysis nudging in the interior of the outermost grid of a nested simulation has been shown to improve the model predictions of the innermost grid (e.g., Wootten et al., 2016). In order to prevent the reflection of waves near the model top, a sponge layer is added (Skamarock et al., 2008). In the vertical, 40 levels are considered, with roughly 17 placed in the lowest 1 km. The hourly output of the 3-km grid in the 20-year simulation and of the 1-km grid in the 1-year run are stored and subsequently used for analysis.

2.2. Observational data

Daily station data from the FMI (available online at <https://en.ilmatieteenlaitos.fi/download-observations/>) are used both for training the CDF-t technique and model evaluation. These data have previously been subjected to quality control, with any unreliable measurements excluded. The daily mean temperature is defined as the average temperature based on four or eight observations per day (0000–2300 UTC). The daily maximum temperature is the highest temperature during the 24-hour period from 2000 Local Standard Time (LST) in the previous evening to 2000 LST in the current evening (2100–2100 LST during summertime). The daily minimum temperature is the lowest temperature observed during the same 24-hour period. In order to allow for a fair comparison with the WRF data, the daily mean, minimum and maximum model-predicted temperatures are computed in the same way as in the FMI data. There are seven stations in the innermost model grids of the two simulations, as shown in Fig. 2, labelled from 1 to 7.

2.3. CDF-t technique

The CDF-t technique, introduced by Michelangeli et al. (2009), has been used in a number of studies, such as Kalache et al. (2011), Colette et al. (2012), Famien et al. (2018) and Wu et al. (2018). For a given variable, the CDF-t relates the CDF obtained from a large-scale model or dataset (e.g., GCM/RCM data), to that observed at a specific location (e.g., weather station data). This technique is developed over a “training” period (e.g., past climate) and ap-

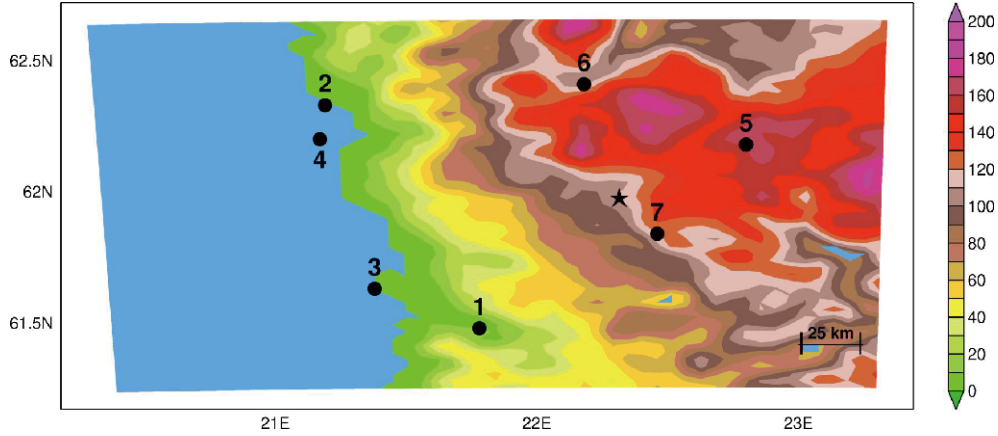


Fig. 2. Spatial distribution of the seven FMI stations (black dots) in the 3-km WRF grid of the 20-year simulation (boundary regions excluded). The shading is the orography (m) as seen by the model and the star highlights the approximate location of the Honkajoki wind farm. The 1-km grid of the 1-year run has a similar spatial extent.

plied over a “prediction” period (e.g., future climate). In the “training” period, the relationship between the large-scale and local-scale’s CDFs is computed. The transformation T between the two is assumed to be time-invariant, so that the local-scale CDF for the “prediction” period can be estimated. T is defined as

$$T[F_{W_p}(x)] = F_{O_p}(x), \quad (1a)$$

where the subscripts W, O and p denote the WRF data, the observed data, and the “training” (past) period, respectively; and x represents the values of the data. Replacing x by $F_{W_p}^{-1}(u)$ in (1a), where $u \in [0, 1]$, we get

$$T(u) = F_{O_p}(F_{W_p}^{-1}(u)). \quad (1b)$$

A similar relation can be obtained for the “prediction” (future) period, f :

$$T[F_{W_f}(x)] = F_{O_f}(x). \quad (2)$$

Substituting Eq. (1b) into Eq. (2) gives

$$F_{O_f}(x) = F_{O_p}[F_{W_p}^{-1}(F_{W_f}(x))]. \quad (3)$$

In Eq. (3), \hat{F}_{O_p} , \hat{F}_{W_p} and \hat{F}_{W_f} can be estimated by the empirical CDFs \hat{F}_{O_p} , \hat{F}_{W_p} and \hat{F}_{W_f} , respectively, so as to estimate \hat{F}_{O_f} , i.e., assuming each data series is stationary ergodic with a common CDF (Dehling and Philipp, 2002). In other words, it is assumed that each data series is stationary, with a common CDF (Dehling and Philipp, 2002). This is a regularly made assumption that, however, has recently been challenged (Dixon et al., 2016). Afterwards, a quantile matching approach is performed between the estimated large-scale CDF $\hat{F}_{W_f}(x)$, and the local-scale CDF $\hat{F}_{O_f}(x)$ to generate the statistically downscaled local climate data (Déqué, 2007; Pierce et al., 2015). In other words, the CDF-t approach first estimates the local-scale CDF for the “prediction” period, and then generates the local-scale climate data,

so that the statistically downscaled data evolve with the large-scale data and the transform T .

Note that $\hat{F}_{O_f}(x)$ is well defined for $x \in [x_{\min}, x_{\max}]$, where x_{\min} and x_{\max} are the minimum and maximum of the future large-scale data, respectively. In some cases, for x outside this interval $\hat{F}_{O_f}(x)$ is constant, either set to $\hat{F}_{O_p}[\hat{F}_{W_p}^{-1}(0)] > 0$ or $\hat{F}_{O_p}[\hat{F}_{W_p}^{-1}(0)] < 1$. A better way to handle $\hat{F}_{O_f}(x)$ for those values of x is to employ the constant correction method suggested by Déqué (2007). It can be summarized as follows: suppose that the percentile $\hat{F}_{O_f}(x_{\min})$ is p . The constant values of $\hat{F}_{O_f}(x)$ for $x < x_{\min}$ are replaced by the CDF of $\hat{F}_{O_p}(x)$ where $\hat{F}_{O_p}(x) < p$. A similar approach is followed for $x > x_{\max}$. In other words, this constant correction approach sets the CDF of a given random variable for the future period for values outside the predicted range for the future but inside the range for the past, to be identical to that of the past period for the corresponding percentile. While the constant correction method has been applied in numerous studies, it has recently been found not to give an optimal treatment of the tails of the distribution (e.g., Lantante et al., 2018, 2019). However, and as highlighted in section 3, this correction is only applied in a very limited number of cases in this work and employing another method will not change the overall findings.

As stated before, the “training” period is taken to be the 20-year (April 1997 to March 2016) period, and the “prediction” period is the central month of each season for April 2016 to March 2017 (i.e., April, July, October 2016 and January 2017). Given the large annual variability of the air temperature in Scandinavia, a transform T is computed for each season separately. For the CDF-t application, the output of the 3-km WRF data is used. The WRF[3 km]+CDF-t predicted temperature and the 1-km WRF-downscaled data are then evaluated against the observed (FMI) air temperature. The model-predicted temperature at the location of a weather station is that given by the closest model grid point to the location of the station to which a topographic correction is ad-

ded. The latter is estimated using the dry adiabatic lapse rate and the difference in elevation between the height of the station and that of the topography of the reference model grid point.

2.4. Verification diagnostics

In order to evaluate the performance of the CDF-t approach, the empirical CDFs of the hybrid statistical–dynamical downscaled WRF[3 km]+CDF-t and the dynamically downscaled WRF[1 km] are compared with the CDF of the observed (FMI) data.

A straightforward way to compare two probability distributions is to plot their CDFs against each other, constructing a Q-Q plot. If the two CDFs are identical, the respective quantiles will be aligned along the main diagonal of the plot (i.e., $y=x$). If they are linearly related but do not match, the Q-Q plot will be a straight line other than the main diagonal. If, on the other hand, the general trend is flatter (steeper) than the $y=x$ line, the distribution plotted on the horizontal axis is more dispersed (clustered) than the distribution plotted on the vertical axis. More details regarding the interpretation of Q-Q plots can be found in Field (2013).

In addition to a visual evaluation, three diagnostics are used to assess the model performance: the mean absolute error (MAE) and the root-mean-square error (RMSE; Wilks, 2006), and the Cramér-von Mises (CvM; Cramer, 1928) criterion.

The MAE between FMI observations and modelled data in the prediction period is given by

$$\text{MAE} = \frac{1}{N_t} \sum_{i=1}^{N_t} |\text{FMI}_i - \text{WRF}_i|, \quad (4)$$

where N_t is the number of points in the period (e.g., $N_t=31$ for July), and the WRF data can be the WRF[3 km]+CDF-t or WRF[1 km] predictions. The RMSE, on the other hand, is expressed as

$$\text{RMSE} = \sqrt{\frac{1}{N_t} \sum_{i=1}^{N_t} (\text{FMI}_i - \text{WRF}_i)^2}. \quad (5)$$

For both diagnostics, the smaller the scores the more skillful the model prediction, and hence the more accurate the method.

A third skill score is used to quantify the overlap between each pair of CDFs. In statistics, for comparing two CDFs, there are three commonly used non-parametric tests: Anderson–Darling (AD), CvM, and Kolmogorov–Smirnov (KS). Each of the three tests has better power against different alternatives; but on the other hand, they exhibit varying degrees of test bias in some situations. Roughly speaking, AD has better power against heavier tails, KS has more power against deviations in the middle, and CvM lies in between (D’Agostino and Stephens, 1986). Hence, CvM is chosen in this work for its balance. The CvM criterion was originally developed to test whether a sample is drawn from a specified continuous distribution and was later generalized by Anderson (1962) to check whether two samples are

drawn from the same unspecified continuous distribution. This diagnostic, CvM, is defined as

$$\text{CvM} = \frac{nm}{n+m} \int_{-\infty}^{+\infty} |\hat{F}_n(x) - \hat{F}_m(x)| d\hat{F}_{n+m}(x), \quad (6)$$

where $\hat{F}_n(x)$ and $\hat{F}_m(x)$ are the empirical CDFs of two samples with sample sizes n and m , respectively, and $\hat{F}_{n+m}(x)$ is the empirical CDF for the combined sample of size $n+m$. For this work, $n=m=30$ for April 2016 and $n=m=31$ for July, October 2016 and January 2017. This criterion can be simplified by using a rank approach. Let r_1, \dots, r_n and s_1, \dots, s_m be the ranks of the two samples in the combined sample, respectively. Anderson (1962) proved that Eq. (6) can be written as

$$\text{CvM} = \frac{U}{nm(n+m)} - \frac{4nm-1}{6(n+m)}; \quad (7)$$

$$U = n \sum_{i=1}^n (r_i - i)^2 + m \sum_{j=1}^m (s_j - j)^2.$$

As the CvM is an increasing function of U given two samples, the CvM as well as U should be larger if the null hypothesis that the two samples come from the same distribution is to be rejected. Thus, the CvM can be seen as a measurement of the divergence between the empirical CDFs of the two samples. In this study, $\hat{F}_m(x)$ is the empirical CDF of the observed (FMI) data, and $\hat{F}_n(x)$ is the empirical CDF to be evaluated with the same time period as $\hat{F}_m(x)$, the CDF of either WRF[3 km]+CDF-t or WRF[1 km] [i.e., $\hat{F}_{\text{W}_f}(x)$]. The smaller the value of the CvM score the better the performance of the downscaling approach, as it will mean that the observed and modelled CDFs have a higher degree of overlap.

After calculating the CvM scores for each station and season, a test is conducted to assess whether the distributions of the scores for all available stations and seasons of the two downscaled products are statistically different. A non-parametric test (i.e., one that does not assume that the data follow a specific distribution), the Wilcoxon rank-sum test (Mann and Whitney, 1947; Fay and Proschan, 2010), is used for this purpose. It compares 28 paired samples, from the seven stations and four seasons, for a given variable (daily mean, maximum and minimum air temperatures). The null hypothesis is generally defined as that it is equally likely that a randomly selected value from sample X will be less or greater than a randomly selected value from another sample Y , i.e. $P(X>Y)=P(X<Y)$. Here, a stronger null hypothesis, namely “the distributions of two samples are equal”, is considered. A downscaling technique is only considered to be more skillful than another if its CvM scores, with respect to the observed data, are statistically significantly lower. More details about this test can be found in Neuhäuser (2011).

3. Results and discussion

3.1. Daily mean temperature

The Q-Q plots for the daily mean air temperature are giv-

en in Fig. 3. Shown are the results for the hybrid downscaling technique (WRF[3 km]+CDF-t) in red and for the higher-resolution dynamic-only downscaling approach (WRF[1 km]) in blue. The numbers at the top of each panel show the MAE between each pair of CDFs, with the panels for a given season sorted out in ascending order of MAE. For all stations and seasons, roughly 1.3% of the total number of points were subjected to the Déqué (2007) out of bounds correction, i.e., less than one point per panel.

Figure 3 shows that the air temperature in the Honkajoki region exhibits a considerable seasonal variability, as expected of a site located at high latitudes. During winter, it hovers around freezing, being warmer than that recorded further east in eastern Finland and neighboring Russia due to the moderating influence of the prevailing midlatitude westerlies, with extreme cold events occurring when the flow is reversed (e.g., Linderson, 2001; Trigo et al., 2002). In the summer, it is lower than that at the referred places, with daily mean values around 290 K, due to the presence of the nearby Gulf of Bothnia and associated mesoscale circulations (e.g., Rummukainen et al., 2001). A visual inspection of Fig. 3 and of the MAE scores reveals that, for the vast majority of the stations and seasons, the hybrid downscaling outperforms the dynamical-only approach. The only stations for which the latter gives the best results are stations 5 and 6 in spring and stations 2, 3 and 5 in winter. While in the spring, summer and autumn seasons the model-predicted temperatures are in general agreement with those observed, in winter the WRF-predictions at all stations exhibit a warm bias, which can exceed 5 K, and has been reported in other studies (e.g., García-Díez et al., 2013). It likely arises from an incorrect simulation of the near-surface atmospheric conditions in stably stratified environments that are ubiquitous in this region and season (Pepin et al., 2009; Steeneveld, 2014). The use of a statistical downscaling technique (CDF-t) does not seem to fix this problem. In spring, on the other hand, in the three coastal stations (2, 3 and 4) WRF has a cold bias, in the three inland stations (5, 6 and 7) it has a warm bias, while at station 1 there is a good agreement between the observed and modelled temperature range. The warm bias in the inland stations may be related to the tendency of the Noah LSM to underestimate the snow cover extent and give a shorter snow season compared to observations (e.g., Sheffield et al., 2003). In fact, and in comparison with the FMI daily data, for the month of April WRF has a tendency to underpredict the snow depth at the location of these stations (not shown). The cold bias in the coastal stations may be due to an incorrect simulation of the observed sea-ice cover/depth over the Gulf of Bothnia, both fields read in from the 6-hourly $0.5^\circ \times 0.5^\circ$ CFSR dataset (Saha et al., 2010), as the sea-ice in the region usually melts in April (Leppäranta and Seinä, 1985). In addition, in the default Noah LSM configuration in WRF there are four soil layers with a maximum depth of 2 m. A lower soil depth may be needed for the model to successfully simulate the freeze–thaw cycles that are important to the Arctic system

(Barlage et al., 2008). For the month of January, the snow depth biases have both positive and negative signs, so they cannot fully explain the warm temperature bias shown in the Q-Q plots in the last column of Fig. 3. For nearly all stations and seasons, the temperature extremes (i.e., the first and last quantiles) are not well captured by both methods. These are harder to simulate, and probably require higher spatial resolutions than those used in this work (3 or 1 km).

The model performance is quantified using the CvM diagnostic, with the scores given in Table 1. As discussed in section 2.4, the smaller the CvM score, the more skillful the model forecast. The results in Table 1 show that, in line with the MAE scores given in Fig. 3, for nearly all stations and seasons the hybrid downscaling approach (WRF[3 km]+CDF-t) gives the smallest CvM values, most of the time by an order of magnitude, compared to those obtained with the dynamical-only downscaling approach (WRF[1 km]). The only exceptions are station 6 in spring and station 2 in winter, for which the hybrid approach gives higher MAEs. Besides an individual inspection of the scores, the Wilcoxon rank-sum test is performed to quantify whether the two distributions of the CvM scores are statistically different. As the CvM scores from the hybrid downscaling approach are generally smaller than the ones from the dynamical-only downscaling approach, a one-sided test is performed, i.e., the alternative hypothesis is that the distribution for the dynamical-only downscaling approach is shifted to the right of the one for the hybrid downscaling approach. The p -value is found to be smaller than 0.05, indicating that there is less than 5% probability that the CvM scores obtained with the two approaches are drawn from the same distribution. In other words, the distributions of the CvM scores are statistically different, with the hybrid approach being closer to the observed data for the daily mean temperature. The CvM scores for the two methods averaged over all stations and seasons are 0.15 and 0.84, respectively. When averaged over all stations and seasons, the RMSE and MAE for the WRF[3 km]+CDF-t are 1.5 K and 1.1 K, whereas for WRF[1 km] they are 1.9 K and 1.6 K, respectively. In other words, all skill scores show that the hybrid method outperforms the dynamical-only downscaling approach, giving predictions generally within 2 K of the observed value. Hence, a WRF[3 km]+CDF-t downscaling technique is not only computationally cheaper than a WRF[1 km] downscaling method, it also gives more skillful predictions of the daily mean air temperature, at least for the Honkajoki region in Finland.

3.2. Daily maximum temperature

In addition to the daily mean temperature, the FMI dataset also provides daily temperature extremes. As they may occur more frequently in a hypothetical warmer world (e.g., Kjellström et al., 2007; Koenig et al., 2013), it is also of interest to consider them. Figure 4 and Table 2 are similar to Fig. 3 and Table 1 but for the daily maximum air temperature. For this field, and for all stations and seasons, roughly 1.5% of the total number of points are subjected to the Déqué (2007) out of bounds correction, i.e., less than one

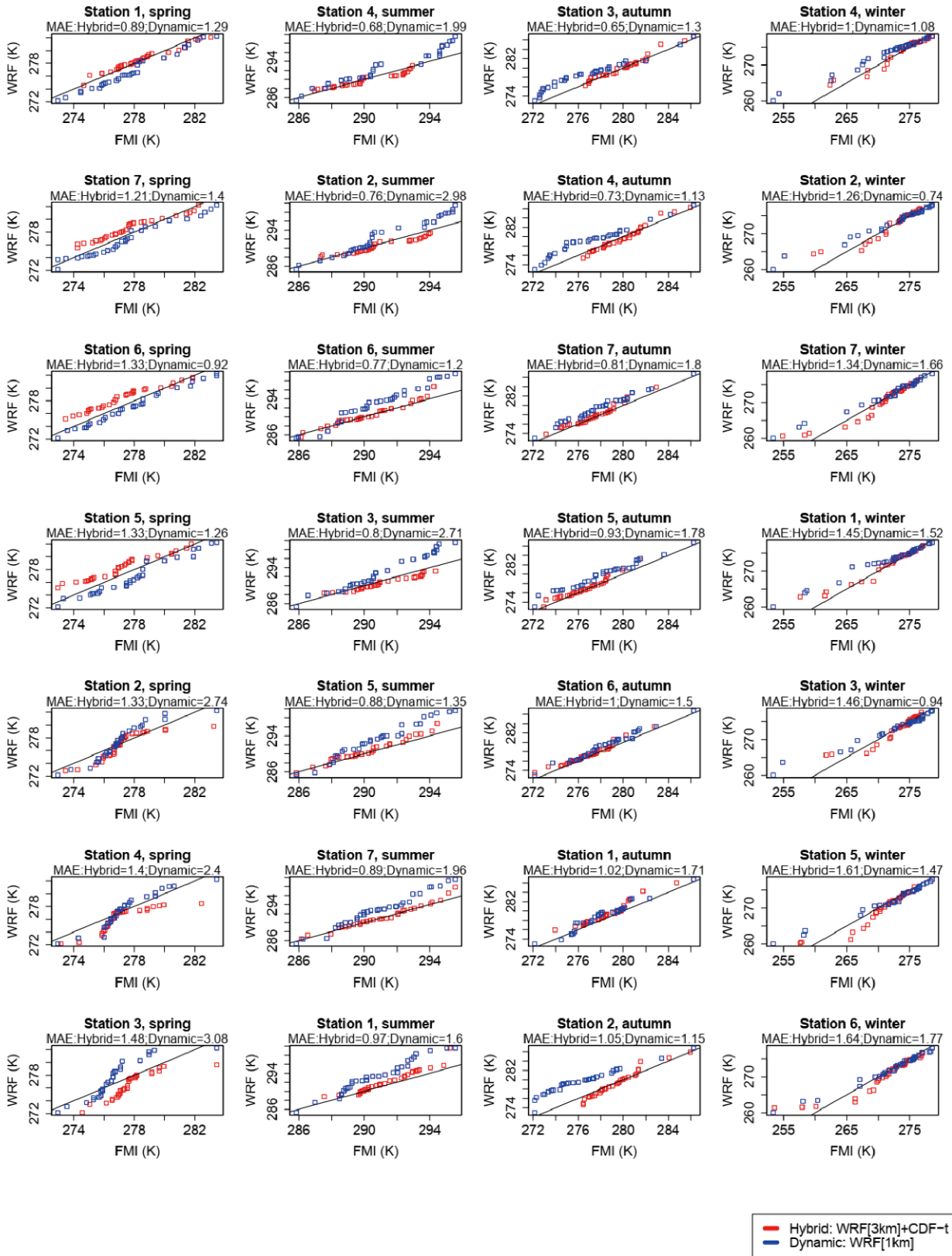


Fig. 3. Q-Q plots for the daily mean air temperature (K), for the central month of each season (April, July, October 2016 and January 2017), and for the seven stations shown in Fig. 2. The red dots represent the 100 quantiles of the dynamically and statistically downscaled (WRF[3km]+CDF-t) data against that observed (FMI). The blue dots are the same but for the higher-resolution dynamically downscaled (WRF[1km]) data. The main diagonal, which indicates a perfect agreement between each pair of CDFs, is drawn as a black line. The numbers at the top of each panel show the MAE (K) between each set of distributions, with the panels for a given season sorted in ascending order of the hybrid method's MAE.

Table 1. CvM scores for the daily mean air temperature obtained with the hybrid product (WRF[3 km]+CDF-t) and the higher-resolution dynamically-only downscaling (WRF[1 km]) with respect to the FMI data, for the central month of each season (April, July, October 2016 and January 2017) and for the seven stations shown in Fig. 2. The CvM scores for the stations/seasons for which (WRF[1 km]) outperforms (WRF[3 km]+CDF-t) are highlighted in bold.

Season	Method	Station						
		1	2	3	4	5	6	7
Spring: April 2016	WRF[3 km]+CDF-t	0.06	0.19	0.54	0.27	0.41	0.48	0.42
	WRF[1 km]	0.72	2.63	3.31	2.38	0.43	0.17	0.56
Summer: July 2016	WRF[3 km]+CDF-t	0.71	0.24	0.29	0.17	0.06	0.05	0.02
	WRF[1 km]	0.91	2.72	2.33	1.40	0.43	0.33	0.95
Autumn: October 2016	WRF[3 km]+CDF-t	0.18	0.08	0.04	0.07	0.05	0.05	0.04
	WRF[1 km]	0.62	0.13	0.20	0.15	0.33	0.22	0.48
Winter: January 2017	WRF[3 km]+CDF-t	0.03	0.07	0.06	0.06	0.07	0.06	0.03
	WRF[1 km]	0.41	0.03	0.30	0.27	0.30	0.37	0.35

point per panel.

As is the case for the daily mean temperature, the hybrid method gives the lowest MAEs for nearly all the stations and seasons, except for station 6 in spring and 2 in winter. In addition, some of the findings reached in the analysis of Fig. 3 hold for the daily maximum temperature, such as (i) warm bias in the model at the location of all stations in winter, at times with a magnitude larger than 5 K; (ii) warm bias in the inland stations (5, 6 and 7) and cold bias in the coastal stations (2, 3 and 4) in spring; (iii) smaller-magnitude biases in summer and autumn, particularly for the WRF[3 km]+CDF-t downscaling; (iv) not so skillful simulation of the extremes of the daily maximum temperature distribution (i.e., the tails of the CDF distributions).

Except for station 2 in winter, the CvM scores for the WRF[3 km]+CDF-t approach are smaller than those obtained with the WRF[1 km] downscaling. For this station, the MAE given by the hybrid approach is also higher. The CvM scores for the two methods averaged over all stations and seasons are 0.14 and 0.91, with RMSE values of 1.8 K and 2.4 K, and MAE values of 1.4 K and 2.1 K, respectively. The p -value obtained with the Wilcoxon rank-sum test with the same alternative hypothesis as for the daily mean temperature is 0.018. Hence, and as is the case for the daily mean temperature, the two distributions of CvM scores are statistically different from each other, with the hybrid approach giving more accurate predictions according to all the verification diagnostics considered. Despite the slightly higher magnitude of the RMSEs and MAEs compared to those of the daily mean temperature, the model-predicted daily maximum temperatures are generally within 2 K of those observed, in particular for the WRF[3 km]+CDF-t downscaling approach. As temperature extremes are likely to change more than the mean values in a hypothetical warmer world (e.g., Seneviratne et al., 2014), it is of interest to look into how the model performs in the warm (summer) and cold (winter) seasons separately. The averaged CvM scores for the hybrid, WRF[3 km]+CDF-t, and dynamical-only, WRF[1 km], downscaling are 0.11 and 1.36 for the summer and 0.13 and 0.47 for the winter season, respectively. The corresponding RMSE (MAE) scores for the hy-

brid and dynamical-only simulations are 1.8 K (1.4 K) and 2.9 K (2.6 K) for the summer, and 1.6 K (1.2 K) and 1.9 K (1.5 K) for the winter season, respectively. Hence, while WRF[3 km]+CDF-t clearly outperforms WRF[1 km] in the summer season, in winter the scores are more comparable, even though the hybrid approach still has the edge.

3.3. Daily minimum temperature

Figure 5 is similar to Fig. 3 and Table 3 is similar to Table 1 but for the daily minimum air temperature. For this field, and for all stations and seasons, roughly 0.6% of the total number of points are subjected to the Déqué (2007) out of bounds correction, i.e., less than one point per panel.

While for the daily mean and maximum air temperatures the hybrid method clearly outperforms the dynamical-only downscaling approach, for the daily minimum air temperature the performance of the two techniques is more comparable, with generally higher MAEs. For this field, both methods show a clear warm bias not just in winter but in all seasons, even though it has a larger magnitude in the cold season. These warmer nighttime temperatures may arise from excessive cloud cover and/or deficiencies in the Noah LSM (e.g. Katragkou et al., 2015; Bastin et al., 2018). For summer and autumn, and except for stations 1 and 2 in the latter, the hybrid approach gives the lowest MAEs. In winter and spring, however, the results are mixed, even though in the former the main difference between the two methods is in the tail of the distribution. The magnitude of the MAEs is also higher for this field: while for the daily mean and maximum air temperatures the hybrid method always gives MAEs lower than 2 K, for this method and for the minimum air temperature the MAEs exceed 2 K at more than 40% of the stations. Recent work has shown that the daily minimum temperature is expected to change more significantly than the daily maximum and mean air temperatures in a hypothetical warmer world (e.g., Nikulin et al., 2011). It is possible then that the poorer performance of the hybrid method for this field may also arise from the violation of the stationary assumption made in the development of the statistical technique.

In line with the Q-Q plots shown in Fig. 5, no method is found to outperform another when inspecting the CvM

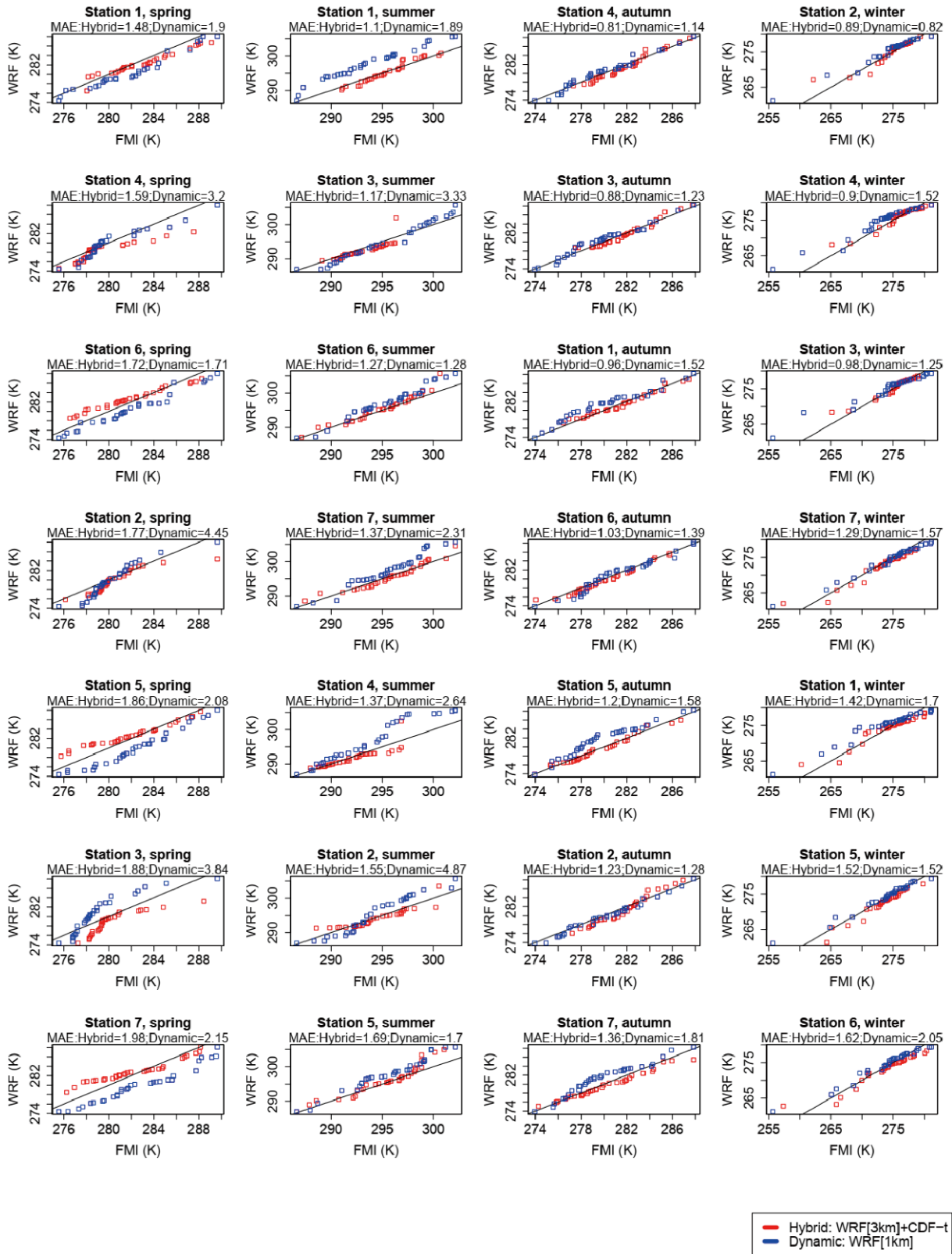


Fig. 4. As in Fig. 3 but for the daily maximum air temperature (K).

scores given in Table 3. A comparison of Table 3 with Tables 1 and 2 reveals that, while for the daily mean and maximum temperatures WRF[1 km] only outperforms WRF[3 km]+CDF-t in two cases and one case, respectively, for

the daily minimum temperature, the dynamical-only downscaling outperforms the hybrid approach in a total of eight cases. Looking at each season separately, four out of the eight cases occur in spring, three in autumn, and one in

Table 2. As Table 1 but for the daily maximum air temperature.

Season	Method	Station						
		1	2	3	4	5	6	7
Spring: April 2016	WRF[3 km]+CDF-t	0.13	0.27	0.49	0.16	0.17	0.21	0.31
	WRF[1 km]	0.73	3.68	3.28	2.64	0.50	0.25	0.37
Summer: July 2016	WRF[3 km]+CDF-t	0.07	0.16	0.30	0.10	0.06	0.04	0.05
	WRF[1 km]	0.40	3.63	2.43	1.61	0.45	0.22	0.78
Autumn: October 2016	WRF[3 km]+CDF-t	0.04	0.17	0.07	0.11	0.04	0.07	0.06
	WRF[1 km]	0.12	0.25	0.20	0.15	0.08	0.11	0.14
Winter: January 2017	WRF[3 km]+CDF-t	0.14	0.21	0.10	0.13	0.16	0.09	0.09
	WRF[1 km]	0.63	0.09	0.31	0.67	0.51	0.69	0.42

winter. The poorer performance of WRF[3 km]+CDF-t seems to be almost exclusively in the transition seasons. As stated before, the WRF forecasts at this time of the year, and particularly in spring, are less skillful, possibly because of deficiencies inherent to the Noah LSM that lead to discrepancies between the observed and modelled snow cover and extent of the snow season (Sheffield et al., 2003). This is confirmed when the observed and modelled snow depths are compared (not shown). The fact that the two methods give more comparable results may suggest that the CDF-t does not work so well in the transition seasons. One possible explanation is that, while in both April and October the daylight period is still relatively long, allowing for a more well-mixed atmosphere and therefore more predictable daytime temperatures, at night the better representation of the static fields and local-scale dynamics in the WRF[1 km] run may give it the edge at the location of some of the stations. Despite this, however, the p -value obtained with the Wilcoxon rank-sum test with the same alternative hypothesis as for daily mean and maximum temperatures is 0.018, indicating that at the 95% confidence level the hybrid downscaling still gives the best results. The averaged CvM scores over all stations and seasons for the hybrid and dynamical-only downscaling methods are 0.33 and 0.64, respectively. For the summer season, the scores for the WRF[3 km]+CDF-t and WRF[1 km] are 0.30 and 1.15, and for the winter season they are 0.08 and 0.22. The corresponding RMSEs (MAEs) values are 2.3 K (1.7 K) and 2.2 K (1.9 K) for all seasons, 1.8 K (1.5 K) and 2.5 K (2.2 K) for the summer season, and 2.9 K (2 K) and 2.1 K (1.7 K) for the winter season, respectively. It is interesting to note that, while for the summer season WRF[3 km]+CDF-t gives the smallest RMSE and MAE, for winter WRF[1 km] performs the best according to those two scores, even though its CvM score is higher. This apparent contradiction can be explained by looking at the Q-Q plots in Fig. 5. The main difference between the two modelled CDFs is in the tail of the distribution, with the CDF of the hybrid method showing larger discrepancies with respect to that observed. However, overall the hybrid method's CDF is closer to the main diagonal compared to that given by the WRF[1 km] simulation, and hence it has a lower CvM but a higher RMSE and MAE. As highlighted before, the model-predicted daily minimum air temperatures forecasts are not

as skillful as those for the daily mean and maximum air temperatures but are still generally within 3 K of those observed.

In the discussion so far the temperature distributions predicted by WRF[3 km]+CDF-t and WRF[1 km] are evaluated against those observed through the analysis of the correspondent Q-Q plots and the MAE and CvM scores. In order to allow for a more direct evaluation of the three temperature distributions, in Fig. 6 they are shown at the location of station 2 for the summer season. Similar results are obtained at the other stations for this season (not shown). The curves plotted in Fig. 6 are in line with the findings highlighted before: WRF[3 km]+CDF-t gives more skillful air temperature predictions compared to WRF[1 km], particularly for the daily-mean and maximum air temperatures. For these two variables, the three temperature distributions are bimodal, with the WRF[1 km] distribution exhibiting a clear cold bias, more significant for the maximum temperature. The two model-predicted distributions are more similar for the daily minimum temperature, being close to a Gaussian shape, as is the case for the observed data, which is consistent with the more comparable MAE and CvM scores given in Fig. 5 and Table 3.

4. Conclusions

The lower spatial resolution and inherent biases of GCM outputs, when evaluated against observational data (e.g., Vigaud et al., 2013), have led to numerous attempts to generate more reliable local-scale predictions using the coarse-resolution model data as input. Both dynamical (using a high-resolution numerical model) and statistical (using a numerical algorithm) downscaling approaches have been conducted earlier with good results (e.g., Warrach-Sagie et al., 2013; Pierce et al., 2015). Hybrid statistical–dynamical downscaling methods, where a statistical technique is applied to numerical model outputs, have also been successfully implemented (e.g., Colette et al., 2012; Famien et al., 2018; Wu et al., 2018). An easy-to-apply, commonly used and computationally less demanding statistical method is the CDF-t (Michelangeli et al., 2009; Pierce et al., 2015). A hybrid statistical–dynamical downscaling method featuring the CDF-t and the WRF model has shown very promising res-

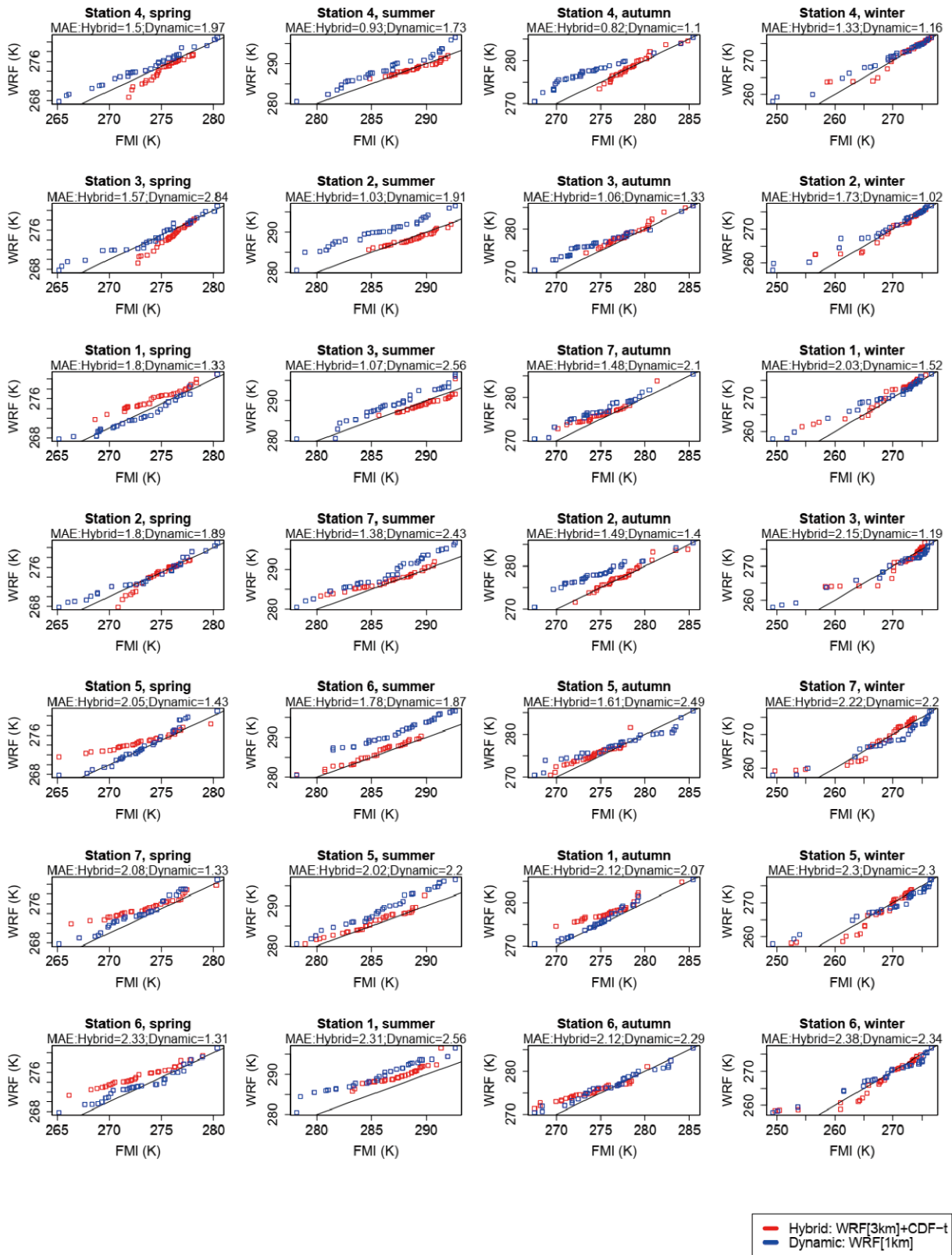


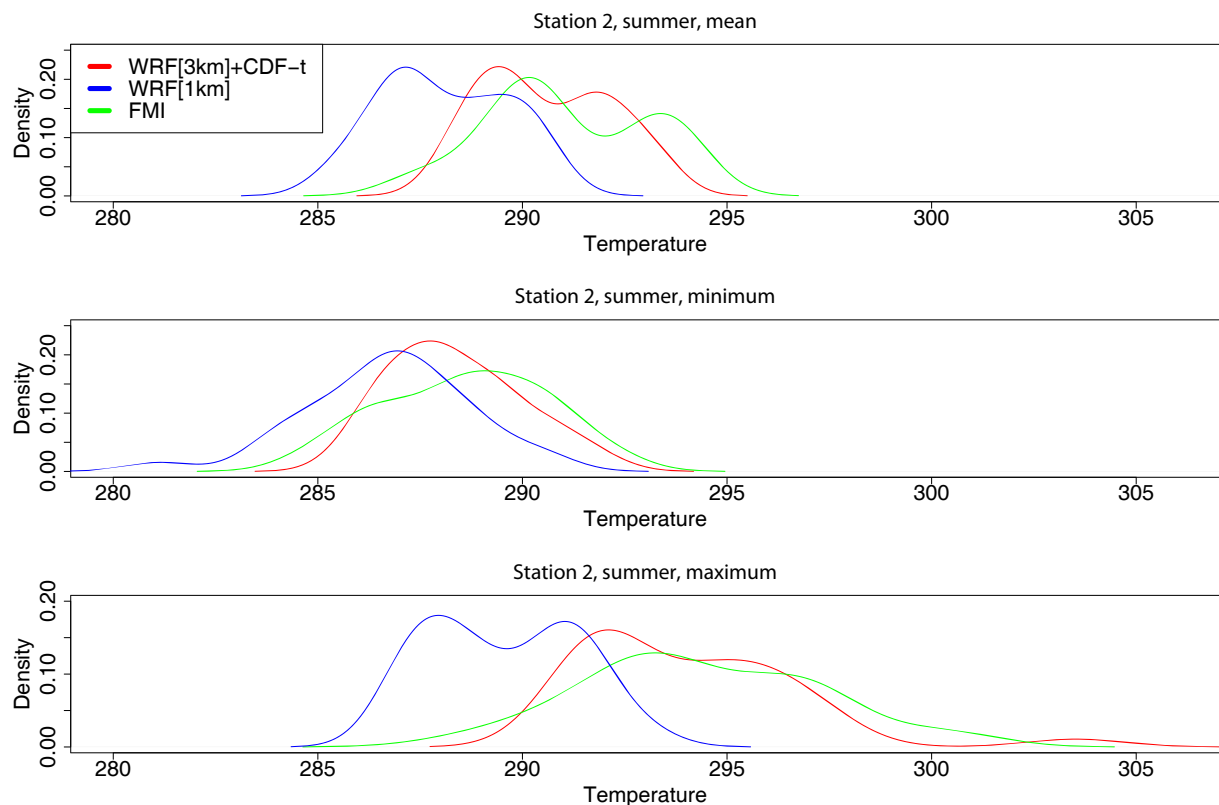
Fig. 5. As in Fig. 3 but for the daily minimum air temperature (K).

ults in simulating local-scale fields that can subsequently be used for climate impact studies (e.g., Colette et al., 2012; Bechler et al., 2015). In this work, the CDF-t is combined with WRF forecasts over the Honkajoki region in Finland,

and evaluated against a higher-resolution dynamical-only downscaling product and observed station data provided by the FMI. The main goal is to assess whether the CDF-t, when applied to a dynamical downscaling of a reanalysis data-

Table 3. As Table 1 but for the daily minimum air temperature.

Season	Method	Station						
		1	2	3	4	5	6	7
Spring: April 2016	WRF[3 km]+CDF-t	0.53	0.06	0.47	0.18	0.89	0.98	0.66
	WRF[1 km]	0.24	1.10	2.44	1.56	0.25	0.10	0.18
Summer: July 2016	WRF[3 km]+CDF-t	1.43	0.12	0.25	0.07	0.09	0.08	0.12
	WRF[1 km]	1.56	1.14	2.14	1.10	0.49	0.59	1.04
Autumn: October 2016	WRF[3 km]+CDF-t	1.14	0.03	0.21	0.04	0.32	0.58	0.56
	WRF[1 km]	0.64	0.05	0.13	0.04	0.80	0.10	0.74
Winter: January 2017	WRF[3 km]+CDF-t	0.18	0.04	0.08	0.02	0.05	0.06	0.10
	WRF[1 km]	0.26	0.03	0.17	0.09	0.33	0.26	0.37

**Fig. 6.** Daily mean, minimum and maximum air temperature distributions (K) from WRF[3 km]+CDF-t (red curve), WRF[1 km] (blue curve) and that observed as given by the FMI data (green curve) at the location of station 2 and for the summer season.

set to a spatial resolution of a few kilometers, can outperform a much more computationally expensive dynamical downscaling to a kilometer resolution. The Honkajoki area is chosen given the large seasonal contrast of the weather conditions in the region, and the fact that RCMs have been shown to have large biases here (e.g., Kotlarski et al., 2014; Katragkou et al., 2015), raising the question of how much of an added value a hybrid downscaling technique will give. The focus of the analysis is on the daily mean and extreme air temperatures, variables that are very relevant for many weather and climate applications, and that are predicted to change significantly in the region in a hypothetical warming climate (e.g., Jylha et al., 2004; Hanssen-Bauer et al., 2005).

For the “training” period, a 20-year (April 1997 to March 2016) WRF downscaling at a 3-km resolution, WRF[3 km], and station data from the FMI are used. A separate transformation T is generated for each season and variable, and then used to predict the daily mean and extremes of air temperature for the central month of each season for the period April 2016 to March 2017. These predictions are then compared to those given by a 1-km WRF downscaling, WRF[1 km]. The computational cost of the WRF[1 km] simulations is the reason why the prediction period is restricted to the central month of each season for one year. This is a limitation of the study, given that the atmospheric conditions in the region are known to exhibit significant variability on interannual time scales (e.g., Toniazio and Scaife, 2006; Gray et al.,

2016). A comprehensive evaluation will be presented in a subsequent paper. The performance of WRF[3 km]+CDF-t is evaluated against that given by WRF[1 km] and the FMI data both visually, through inspection of the correspondent Q-Q plots, and quantitatively, using the CvM, RMSE and MAE diagnostics. It is concluded that, at the 95% confidence level, the hybrid approach, particularly for the daily mean and maximum air temperatures, gives more skillful forecasts compared to the dynamical-only downscaling at a higher spatial resolution, at least for the near-coast Finnish region studied. For the daily minimum air temperature, however, the two methods perform comparably.

For all stations and variables, WRF has a warm bias in winter that has been noted by García-Díez et al. (2013) and is possibly due to deficiencies in the simulation of stably stratified environments, a regular occurrence at high latitudes in the cold season. In spring, on the other hand, WRF yields temperatures that are too warm at the inland stations and too cold in coastal areas. While the former may be related to the tendency of the Noah LSM to underpredict the amount of snow cover and the length of the snow season (e.g., Sheffield et al., 2003), the latter may arise from an incorrect representation of the sea-ice cover and extent, perhaps the timing of the spring melt, that are read in from the 6-hourly CFSR data. The daily minimum temperatures in WRF are too warm in all seasons, not just in winter, although the biases have a larger magnitude in the cold season. A possible explanation is excessive cloud cover and/or potential deficiencies in the Noah LSM (e.g., Katragkou et al., 2015; Bastin et al., 2018).

For the daily mean and maximum temperatures, the hybrid approach, WRF[3 km]+CDF-t, gives the best scores at nearly all stations and seasons, with averaged CvM values of about 0.15 and 0.14, RMSE values of 1.5 K and 1.8 K, and MAE values of 1.9 K and 1.4 K, respectively. On the other hand, for the daily minimum temperature, the results obtained with the two techniques are more comparable, with WRF[1 km] outperforming WRF[3 km]+CDF-t for four stations in spring, three in autumn, and one in winter. In addition to the more variable weather conditions in the transition seasons that may be harder to simulate with the CDF-t, the predicted more significant changes in the daily minimum air temperature compared to the daily maximum and mean temperatures in a future warming climate (e.g., Nikulin et al., 2011) may suggest that the stationary assumption made in the CDF-t technique does not work so well for this field. In any case, and for the hybrid approach, the CvM score for the daily minimum temperature is about 0.33, the RMSE is 2.3 K, and the MAE is 1.7 K. These scores show an overall good performance and highlight the potential use of this WRF product for in-situ air temperature estimation, at least for this region. These results are confirmed by a direct inspection of the observed, WRF[3 km]+CDF-t and WRF[1 km] air temperature distributions.

As the CDF-t does not require much time and computational resource to implement, this methodology can easily be extended to other variables (e.g., precipitation, and quantit-

ies related to wind energy and icing applications), regions, numerical models and future climate change studies. Some of the referred extensions of the work will be presented in a subsequent paper. Future work will also test better treatments of the tail adjustment of the modelled CDF, such as the one proposed by Lanzante et al. (2019).

Acknowledgements. We acknowledge Botnia-Atlantica, an EU-programme financing cross border cooperation projects in Sweden, Finland and Norway, for their support of this work through the WindCoE project. We would like to thank the High Performance Computing Center North for providing the computer resources needed to perform the numerical experiments presented in this paper. We would like to thank the three anonymous reviewers for their detailed and insightful comments and suggestions, which helped to improve the quality of the paper.

REFERENCES

- Aitken, M. L., B. Kosović, J. D. Mirocha, and J. K. Lundquist, 2014: Large eddy simulation of wind turbine wake dynamics in the stable boundary layer using the Weather Research and Forecasting Model. *Journal of Renewable and Sustainable Energy*, **6**, 033137, <https://doi.org/10.1063/1.4885111>.
- Alexandru, A., R. de Elia, R. Laprise, L. Separovic, and S. Biner, 2009: Sensitivity study of regional climate model simulations to large-scale nudging parameters. *Mon. Wea. Rev.*, **137**, 1666–1686, <https://doi.org/10.1175/2008MWR2620.1>.
- Anderson, T. W., 1962: On the distribution of the two-sample cramer-von mises criterion. *The Annals of Mathematical Statistics*, **33**, 1148–1159, <https://doi.org/10.1214/aoms/1177704477>.
- Barlage, M., F. Chen, and D. Bromwich, 2008: Noah Land Surface Model Soil Depth Modifications for Arctic Regional Climate Simulations. *American Geophysical Union Fall Meeting 2008*, San Francisco, CA, American Geophysical Union. [Available online from https://ral.ucar.edu/~barlage/ASR/presentations/Barlage_AGU_Poster_12Dec08.pdf.]
- Bastin, S., M. Chiriac, and P. Drobinski, 2018: Control of radiation and evaporation on temperature variability in a WRF regional climate simulation: Comparison with colocated long term ground based observations near Paris. *Climate Dyn.*, **51**, 985–1003, <https://doi.org/10.1007/s00382-016-2974-1>.
- Bechler, A., M. Vrac, and L. Bel, 2015: A spatial hybrid approach for downscaling of extreme precipitation fields. *J. Geophys. Res.*, **120**, 4534–4550, <https://doi.org/10.1002/2014JD022558>.
- Bowden, J. H., T. L. Otte, C. G. Nolte, and M. J. Otte, 2012: Examining interior grid nudging techniques using two-way nesting in the WRF model for regional climate modeling. *J. Climate*, **25**, 2805–2823, <https://doi.org/10.1175/JCLI-D-11-00167.1>.
- Bowden, J. H., C. G. Nolte, and T. L. Otte, 2013: Simulating the impact of the large-scale circulation on the 2-m temperature and precipitation climatology. *Climate Dyn.*, **40**, 1903–1920, <https://doi.org/10.1007/s00382-012-1440-y>.
- Bruyère, C. L., J. M. Done, G. J. Holland, and S. Fredrick, 2014: Bias corrections of global models for regional climate simulations of high-impact weather. *Climate Dyn.*, **43**, 1847–1856, <https://doi.org/10.1007/s00382-013-2011-6>.

- Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.*, **129**, 569–585, [https://doi.org/10.1175/1520-0493\(2001\)129<0569:CAALSH>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2).
- Colette, A., R. Vautard, and M. Vrac, 2012: Regional climate downscaling with prior statistical correction of the global climate forcing. *Geophys. Res. Lett.*, **39**, L13707, <https://doi.org/10.1029/2012GL052258>.
- Collins, W. D., and Coauthors, 2006: The community climate system model version 3 (CCSM3). *J. Climate*, **19**, 2122–2143, <https://doi.org/10.1175/JCLI3761.1>.
- Cramer, H., 1928: On the composition of elementary errors. *Scandinavian Actuarial Journal*, **1928**, 141–180, <https://doi.org/10.1080/03461238.1928.10416872>.
- Davis, N., A. N. Hahmann, N.-L. Clausen, and M. Žagar, 2014: Forecast of icing events at a wind farm in Sweden. *Journal of Applied Meteorology and Climatology*, **53**, 262–281, <https://doi.org/10.1175/JAMC-D-13-09.1>.
- D'Agostino, R. B., and M. A. Stephens, 1986: *Goodness-of-Fit Techniques*. Marcel Dekker, Inc.
- Dehling, H., and W. Philipp, 2002: Empirical process techniques for dependent data. *Empirical Process Techniques for Dependent Data*, H. G. Dehling, T. Mikosch, and M. Sørensen, Eds., Birkhäuser, 3–113, https://doi.org/10.1007/978-1-4612-0099-4_1.
- Déqué, M., 2007: Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values. *Global and Planetary Change*, **57**, 16–26, <https://doi.org/10.1016/j.gloplacha.2006.11.030>.
- Diaz-Nieto, J., and R. L. Wilby, 2005: A comparison of statistical downscaling and climate change factor methods: Impacts on low flows in the River Thames, United Kingdom. *Climatic Change*, **69**, 245–268, <https://doi.org/10.1007/s10584-005-1157-6>.
- Dixon, K. W., J. R. Lanzante, M. J. Nath, K. Hayhoe, A. Stoner, A. Radhakrishnan, V. Balaji, and C. F. Gaitan, 2016: Evaluating the stationarity assumption in statistically downscaled climate projections: Is past performance an indicator of future results? *Climatic Change*, **135**, 395–408, <https://doi.org/10.1007/s10584-016-1598-0>.
- Dulière, V., Y. X. Zhang, and E. P. Salathé Jr., 2011: Extreme precipitation and temperature over the U.S. Pacific Northwest: A comparison between observations, reanalysis data, and regional models. *J. Climate*, **24**, 1950–1964, <https://doi.org/10.1175/2010JCLI3224.1>.
- Duran, P., C. Meissner, K. Rutledge, R. Fonseca, J. Martin-Torres, and M. S. Adaramola, 2019: Meso-microscale coupling for wind resource assessment using averaged atmospheric stability conditions. *Meteorol. Z.*, <https://doi.org/10.1127/metz/2019/0937>.
- Duynkerke, P. G., 1991: Radiation fog: A comparison of model simulation with detailed observations. *Mon. Wea. Rev.*, **119**, 324–341, [https://doi.org/10.1175/1520-0493\(1991\)119<0324:RFACOM>2.0.CO;2](https://doi.org/10.1175/1520-0493(1991)119<0324:RFACOM>2.0.CO;2).
- Ebisuzaki, W., and L. Zhang, 2011: Assessing the performance of the CFSR by an ensemble of analyses. *Climate Dyn.*, **37**, 2541–2550, <https://doi.org/10.1007/s00382-011-1074-5>.
- Famien, A. M., S. Janicot, A. D. Ochou, M. Vrac, D. Defrance, B. Sultan, and T. Noël, 2018: A bias-corrected CMIP5 dataset for Africa using the CDF-t method—a contribution to agricultural impact studies. *Earth System Dynamics*, **9**, 313–338, <https://doi.org/10.5194/esd-9-313-2018>.
- Fan, K., Y. Liu, and H. P. Chen, 2012: Improving the prediction of the East Asian summer monsoon: New approaches. *Wea. Forecasting*, **27**, 1017–1030, <https://doi.org/10.1175/WAF-D-11-00092.1>.
- Fay, M. P., and M. A. Proschan, 2010: Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Statistics Surveys*, **4**, 1–39, <https://doi.org/10.1214/09-SS051>.
- Field, A. P., 2013: *Discovering Statistics Using IBM SPSS Statistics: and Sex and Drugs and Rock 'N' Roll*. 4th ed. SAGE, 915 pp.
- Fitch, A. C., J. B. Olson, J. K. Lundquist, J. Dudhia, A. K. Gupta, J. Michalak, and I. Barstad, 2012: Local and mesoscale impacts of wind farms as parameterized in a mesoscale NWP model. *Mon. Wea. Rev.*, **140**, 3017–3038, <https://doi.org/10.1175/MWR-D-11-00352.1>.
- Flaounas, E., P. Drobinski, M. Vrac, S. Bastin, C. Lebeaupin-Brossier, M. Stéfanon, M. Borga, and J.-C. Calvet, 2013: Precipitation and temperature space-time variability and extremes in the Mediterranean region: Evaluation of dynamical and statistical downscaling methods. *Climate Dyn.*, **40**, 2687–2705, <https://doi.org/10.1007/s00382-012-1558-y>.
- Fowler, H. J., S. Blenkinsop, and C. Tebaldi, 2007: Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology*, **27**, 1547–1578, <https://doi.org/10.1002/joc.1556>.
- Free, M., B. M. Sun, and H. L. Yoo, 2016: Comparison between total cloud cover in four reanalysis products and cloud measured by visual observations at U.S. weather stations. *J. Climate*, **29**, 2015–2021, <https://doi.org/10.1175/JCLI-D-15-0637.1>.
- García-Díez, M., J. Fernández, L. Fita, and C. Yagüe, 2013: Seasonal dependence of WRF model biases and sensitivity to PBL schemes over Europe. *Quart. J. Roy. Meteor. Soc.*, **139**, 501–514, <https://doi.org/10.1002/qj.1976>.
- Gray, L. J., T. J. Woollings, M. Andrews, and J. Knight, 2016: Eleven-year solar cycle signal in the NAO and Atlantic/European blocking. *Quart. J. Roy. Meteor. Soc.*, **142**, 1890–1903, <https://doi.org/10.1002/qj.2782>.
- Hanssen-Bauer, I., C. Achberger, R. E. Benestad, D. Chen, and E. J. Forland, 2005: Statistical downscaling of climate scenarios over Scandinavia. *Climate Research*, **29**, 255–268, <https://doi.org/10.3354/cr029255>.
- Heikkilä, U., A. Sandvik, and A. Sorteberg, 2014: Dynamical downscaling of ERA-40 in complex terrain using the WRF regional climate model. *Climate Dyn.*, **37**, 1551–1564, <https://doi.org/10.1007/s00382-010-0928-6>.
- Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, <https://doi.org/10.1029/2008JD009944>.
- Janjić, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945, [https://doi.org/10.1175/1520-0493\(1994\)122<0927:TSMECM>2.0.CO;2](https://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2).

- tps://doi.org/10.1175/1520-0493(1994)122<0927:TS-MECM>2.0.CO;2.
- Jylha, K., H. Tuomenvirta, and K. Ruosteenoja, 2004: Climate change projections for Finland during the 21st century. *Boreal Environment Research*, **9**, 127–152.
- Kallache, M., M. Vrac, P. Naveau, and P. A. Michelangeli, 2011: Nonstationary probabilistic downscaling of extreme precipitation. *J. Geophys. Res.*, **116**, 05113, <https://doi.org/10.1029/2010JD014892>.
- Katata, G., M. Kajino, T. Hiraki, M. Aikawa, T. Kobayashi, and H. Nagai, 2011: A method for simple and accurate estimation of fog deposition in a mountain forest using a meteorological model. *J. Geophys. Res.*, **116**, D20102, <https://doi.org/10.1029/2010JD015552>.
- Katragkou, E., and Coauthors, 2015: Regional climate Hindcast simulations within EURO-CORDEX: Evaluation of a WRF multi-physics ensemble. *Geoscientific Model Development*, **8**, 603–618, <https://doi.org/10.5194/gmd-8-603-2015>.
- Ke, Z. J., P. Q. Zhang, L. J. Chen, and L. M. Du, 2011: An experiment of a statistical downscaling forecast model for summer precipitation over China. *Atmospheric and Oceanic Science Letters*, **4**, 270–275, <https://doi.org/10.1080/16742834.2011.11446941>.
- Kidston, J., A. A. Scaife, S. C. Hardiman, D. M. Mitchell, N. Butchart, M. P. Baldwin, and L. J. Gray, 2015: Stratospheric influence on tropospheric jet streams, storm tracks and surface weather. *Nature Geoscience*, **6**, 433–440, <https://doi.org/10.1038/NGEO2424>.
- Kjellström, E., L. Bärring, D. Jacob, R. Jones, G. Lenderink, and C. Schär, 2007: Modelling daily temperature extremes: Recent climate and future changes over Europe. *Climatic Change*, **81**, 249–265, <https://doi.org/10.1007/s10584-006-9220-5>.
- Koenig, T., L. Brodeau, R. G. Graversen, J. Karlsson, G. Svensson, M. Tjernström, U. Willén, and K. Wyser, 2013: Arctic climate change in 21st century CMIP5 simulations with EC-Earth. *Climate Dyn.*, **40**, 2719–2743, <https://doi.org/10.1007/s00382-012-1505-y>.
- Koh, T.-Y., and R. M. Fonseca, 2016: Subgrid-scale cloud-radiation feedback for the Betts-miller- Janjić convection scheme. *Quart. J. Roy. Meteor. Soc.*, **142**, 989–1006, <https://doi.org/10.1002/qj.2702>.
- Kotlarski, S., and Coauthors, 2014: Regional climate modeling on European scales: A joint standard evaluation of the EURO-CORDEX RCM ensemble. *Geoscientific Model Development*, **7**, 1297–1333, <https://doi.org/10.5194/gmd-7-1297-2014>.
- Landman, W. A., and W. J. Tennant, 2000: Statistical downscaling of monthly forecasts. *International Journal of Climatology*, **20**, 1521–1532, [https://doi.org/10.1002/1097-0088\(20001115\)20:13<1521::AID-JOC558>3.0.CO;2-N](https://doi.org/10.1002/1097-0088(20001115)20:13<1521::AID-JOC558>3.0.CO;2-N).
- Lanzante, J. R., K. W. Dixon, M. J. Nath, C. E. Whitlock, and D. Adams-Smith, 2018: Some pitfalls in statistical downscaling of future climate. *Bull. Amer. Meteor. Soc.*, **99**, 791–803, <https://doi.org/10.1175/BAMS-D-17-0046.1>.
- Lanzante, J. R., M. J. Nath, C. E. Whitlock, K. W. Dixon, and D. Adams-Smith, 2019: Evaluation and improvement of tail behaviour in the cumulative distribution function transform downscaling method. *International Journal of Climatology*, **39**, 2449–2460, <https://doi.org/10.1002/joc.5964>.
- Laprise, R., 2008: Regional climate modelling. *J. Comput. Phys.*, **227**, 3641–3666, <https://doi.org/10.1016/j.jcp.2006.10.024>.
- Lavaysse, C., M. Vrac, P. Drobinski, M. Lengaigne, and T. Vichet, 2012: Statistical downscaling of the French Mediterranean climate: Assessment for present and projection in an anthropogenic scenario. *Natural Hazards and Earth System Sciences*, **12**, 651–670, <https://doi.org/10.5194/nhess-12-651-2012>.
- Leppäranta, M., and A. Seinä, 1985: Freezing, maximum annual ice thickness and breakup of ice on the Finnish coast during 1830–1984. *Geophysica*, **21**, 87–104.
- Linderson, M.-J., 2001: Objective classification of atmospheric circulation over southern Scandinavia. *International Journal of Climatology*, **21**, 155–169, <https://doi.org/10.1002/joc.604>.
- Lo, J. C. F., Z. L. Yang, and R. A. Pielke, 2008: Assessment of three dynamical climate downscaling methods using the Weather Research and Forecasting (WRF) Model. *Geophys. Res. Lett.*, **113**, D09112, <https://doi.org/10.1029/2007JD009216>.
- Ma, Y. Y., Y. Yang, X. P. Mai, C. J. Qiu, X. Long, and C. H. Wang, 2016: Comparison of analysis and spectral nudging techniques for dynamical downscaling with the WRF model over China. *Advances in Meteorology*, **2016**, 4761513, <https://doi.org/10.1155/2016/4761513>.
- Mann, H. B., and D. R. Whitney, 1947: On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, **18**, 50–60, <https://doi.org/10.1214/aoms/1177730491>.
- Merino, A., and Coauthors, 2015: Cloud top height estimation from WRF model: Application to the infrared camera on-board EUSO-Balloon (CNES). *Proceedings of the 34th International Cosmic Ray Conference*, 30th July–6th August, The Hague, The Netherlands.
- Michelangeli, P.-A., M. Vrac, and H. Loukos, 2009: Probabilistic downscaling approaches: Application to wind cumulative distribution functions. *Geophys. Res. Lett.*, **36**, L11708, <https://doi.org/10.1029/2009GL038401>.
- Miguez-Macho, G., G. L. Stenchikov, and A. Robock, 2004: Spectral nudging to eliminate the effects of domain position and geometry in regional climate model simulations. *J. Geophys. Res.*, **109**, D13104, <https://doi.org/10.1029/2003JD004495>.
- Miguez-Macho, G., G. L. Stenchikov, and A. Robock, 2005: Regional climate simulations over North America: Interaction of local processes with improved large-scale flow. *J. Climate*, **18**, 1227–1246, <https://doi.org/10.1175/JCLI3369.1>.
- Mills, C. M., 2011: On the weather research and forecasting model's treatment of sea ice albedo over the arctic ocean. *Proc. 10th Annual School of Earth, Society, and Environmental Research Review*, Urbana-Campaign, IL, University of Illinois.
- Monin, A. S., and A. M. Obukhov, 1954: Basic laws of turbulent mixing in the ground layer of the atmosphere. *Trans. Geophys. Inst. Akad. Nauk USSR*, **151**, 163–187.
- Nakanishi, M., 2000: Large-eddy simulation of radiation fog. *Bound.-Layer Meteor.*, **94**, 461–493, <https://doi.org/10.1023/A:1002490423389>.
- Nakanishi, M., and H. Niino, 2006: An improved Mellor-Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Bound.-Layer Meteor.*, **119**, 397–407, <https://doi.org/10.1007/s10546-005-9030-8>.

- Neuhäuser, M., 2011: Wilcoxon-Mann-Whitney test. *International Encyclopedia of Statistical Science*, M. Lovric, Ed., Springer, https://doi.org/10.1007/978-3-642-04898-2_615.
- Nikulin, G., E. Kjellström, U. Hansson, G. Strandberg, and A. Ullerstig, 2011: Evaluation and future projections of temperature, precipitation and wind extremes over Europe in an ensemble of regional climate simulations. *Tellus A: Dynamic Meteorology and Oceanography*, **63**, 41–55, <https://doi.org/10.1111/j.1600-0870.2010.00466.x>.
- Otte, T. L., C. G. Nolte, M. J. Otte, and J. H. Bowden, 2012: Does nudging squelch the extremes in regional climate modeling? *J. Climate*, **25**, 7046–7066, <https://doi.org/10.1175/JCLI-D-12-00048.1>.
- Overland, J. E., and Coauthors, 2016: Nonlinear response of mid-latitude weather to the Changing Arctic. *Nat. Clim. Change*, **6**, 992–999, <https://doi.org/10.1038/nclimate3121>.
- Pan, X. D., X. Li, X. K. Shi, X. J. Han, L. H. Luo, and L. X. Wang, 2012: Dynamic downscaling of near-surface air temperature at the basin scale using WRF—a case study in the Heihe River Basin, China. *Frontiers of Earth Science*, **6**, 314–323, <https://doi.org/10.1007/s11707-012-0306-2>.
- Pepin, N. C., M. K. Schaefer, and L. D. Riddy, 2009: Quantification of the cold-air pool in Kevo valley, Finnish Lapland. *Weather*, **64**, 60–67, <https://doi.org/10.1002/wea.260>.
- Pierce, D. W., D. R. Cayan, E. P. Maurer, J. T. Abatzoglou, and K. C. Hegewisch, 2015: Improved bias correction techniques for hydrological simulations of climate change. *Journal of Hydrometeorology*, **16**, 2421–2442, <https://doi.org/10.1175/JHM-D-14-0236.1>.
- Rummukainen, M., J. Räisänen, B. Bringfelt, A. Ullerstig, A. Omsstedt, U. Willén, U. Hansson, and C. Jones, 2001: A regional climate model for northern Europe: model description and results from the downscaling of two GCM control simulations. *Climate Dyn.*, **17**, 339–359, <https://doi.org/10.1007/s003820000109>.
- Saha, S., and Coauthors, 2010: The NCEP climate forecast system reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, <https://doi.org/10.1175/2010BAMS3001.1>.
- Seneviratne, S. I., M. G. Donat, B. Mueller, and L. V. Alexander, 2014: No pause in the increase of hot temperature extremes. *Nat. Clim. Change*, **4**, 161–163, <https://doi.org/10.1038/nclimate2145>.
- Sheffield, J., and Coauthors, 2003: Snow process modeling in the North American Land Data Assimilation System (NLDAS): 1. Evaluation of model-simulated snow cover extent. *J. Geophys. Res.*, **108**, GCP 10, <https://doi.org/10.1029/2002JD003274>.
- Skamarock, W. C., and Coauthors, 2008: A description of the advanced research WRF version 3. NCAR Technical Note TN-4175 STR, 113 pp, <https://doi.org/10.5065/D68S4MVH>.
- Soares, P. M. M., R. M. Cardoso, P. M. A. Miranda, J. de Medeiros, M. Belo-Pereira, and F. Espirito-Santo, 2012: WRF high resolution dynamical downscaling of ERA-interim for Portugal. *Climate Dyn.*, **39**, 2497–2522, <https://doi.org/10.1007/s00382-012-1315-2>.
- Stauffer, D. R., and N. L. Seaman, 1990: Use of four-dimensional data assimilation in a limited-area mesoscale model. Part I: Experiments with synoptic-scale data. *Mon. Wea. Rev.*, **118**, 1250–1277, [https://doi.org/10.1175/1520-0493\(1990\)118<1250:UOFDDA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1990)118<1250:UOFDDA>2.0.CO;2).
- Steenefeld, G.-J., 2014: Current challenges in understanding and forecasting stable boundary layers over land and ice. *Frontiers in Environmental Science*, **2**, 41, <https://doi.org/10.3389/fenvs.2014.00041>.
- Steinhoff, D. F., D. H. Bromwich, J. C. Speirs, H. A. McGowan, and A. J. Monaghan, 2014: Austral summer foehn winds over the McMurdo dry valleys of Antarctica from Polar WRF. *Quart. J. Roy. Meteor. Soc.*, **140**, 1825–1837, <https://doi.org/10.1002/qj.2278>.
- Tegen, I., P. Hollrig, M. Chin, I. Fung, D. Jacob, and J. Penner, 1997: Contribution of different aerosol species to the global aerosol extinction optical thickness: Estimates from model results. *J. Geophys. Res.*, **102**, 23 895–23 915, <https://doi.org/10.1029/97JD01864>.
- Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, <https://doi.org/10.1175/2008MWR2387.1>.
- Toniazzo, T., and A. A. Scaife, 2006: The influence of ENSO on winter North Atlantic climate. *Geophys. Res. Lett.*, **33**, L24704, <https://doi.org/10.1029/2006GL027881>.
- Trigo, R. M., T. J. Osborn, and J. M. Corte-Real, 2002: The North Atlantic Oscillation influence on Europe: Climate impacts and associated physical mechanisms. *Climate Research*, **20**, 9–17, <https://doi.org/10.3354/cr020009>.
- Vigaud, N., M. Vrac, and Y. Caballero, 2013: Probabilistic downscaling of GCM scenarios over southern India. *International Journal of Climatology*, **33**, 1248–1263, <https://doi.org/10.1002/joc.3509>.
- von Storch, H., H. Langenberg, and F. Feser, 2000: A spectral nudging technique for dynamical downscaling purposes. *Mon. Wea. Rev.*, **128**, 3664–3673, [https://doi.org/10.1175/1520-0493\(2000\)128<3664:ASNTFD>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<3664:ASNTFD>2.0.CO;2).
- Vrac, M., P. Drobinski, A. Merlo, M. Herrmann, C. Lavaysse, L. Li, and S. Somot, 2012: Dynamical and statistical downscaling of the French Mediterranean climate: uncertainty assessment. *Natural Hazards and Earth System Sciences*, **12**, 2769–2784, <https://doi.org/10.5194/nhess-12-2769-2012>.
- Waldron, K. M., J. Paegle, and J. D. Horel, 1996: Sensitivity of a spectrally filtered and nudged limited-area model to outer model options. *Mon. Wea. Rev.*, **124**, 529–547, [https://doi.org/10.1175/1520-0493\(1996\)124<0529:SOASFA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<0529:SOASFA>2.0.CO;2).
- Wang, J. F., R. M. Fonseca, K. Rutledge, J. Martín-Torres, and J. Yu, 2019: Weather simulation uncertainty estimation using Bayesian hierarchical models. *Journal of Applied Meteorology and Climatology*, **58**, 585–603, <https://doi.org/10.1175/JAMC-D-18-0018.1>.
- Warrach-Sagi, K., T. Schwitalla, V. Wulfmeyer, and H.-S. Bauer, 2013: Evaluation of a climate simulation in Europe based on the WRF-NOAH model system: Precipitation in Germany. *Climate Dyn.*, **41**, 755–774, <https://doi.org/10.1007/s00382-013-1727-7>.
- Wei, F. Y., and J. Y. Huang, 2010: A study of predictability for summer precipitation on East China using downscaling techniques. *Journal of Tropical Meteorology*, **26**, 483–488, <https://doi.org/10.3969/j.issn.1004-4965.2010.04.013>. (in Chinese with English abstract)
- Wilby, R. L., and T. M. L. Wigley, 1997: Downscaling general circulation model output: A review of methods and limitations.

- Progress in Physical Geography: Earth and Environment*, **21**, 530–548, <https://doi.org/10.1177/030913339702100403>.
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Academic Press.
- Wooten, A., J. J. Bowden, R. Boyles, and A. Terando, 2016: The sensitivity of WRF downscaled precipitation in Puerto Rico to cumulus parameterization and interior grid nudging. *Journal of Applied Meteorology and Climatology*, **65**, 2263–2281, <https://doi.org/10.1175/JAMC-D-16-0121.1>.
- Wu, W., Z. R. Liang, and X. C. Liu, 2018: Projection of the daily precipitation using CDF-T method at meteorological observation site scale. *Plateau Meteorology*, **37**, 796–805, <https://doi.org/10.7522/j.issn.1000-0534.2017.00064>. (in Chinese with English abstract)
- Zeng, X. B., and A. Beljaars, 2005: A prognostic scheme of sea surface skin temperature for modeling and data assimilation. *Geophys. Res. Lett.*, **32**, L14605, <https://doi.org/10.1029/2005GL023030>.
- Zorita, E., and H. von Storch, 1999: The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. *J. Climate*, **12**, 2474–2489, [https://doi.org/10.1175/1520-0442\(1999\)012<2474:TAMAAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<2474:TAMAAS>2.0.CO;2).